Clip-level Feature Aggregation for Video-based Person Re-Identification

Chengjin Lyu, Patrick Heyer Wollenberg, Ljiljana Platisa, Bart Goossens, and Wilfried Philips
TELIN-IPI, Ghent University - imec, 9000 Gent, Belgium

Abstract: In the task of video-based person re-identification, sequence-level features of query and gallery persons are compared to search the best matching. Generally, due to the memory limitation of a single GPU, frame-level features are aggregated together using a temporal modeling method to generate clip-level features, instead of a sequence-level representation. Although these clip-level features have achieved impressive results, the importance of clip-level feature aggregation is still lack of study. In this paper, we investigate the aggregation of clip-level features in video-based person re-identification. Specifically, a simple but useful clip-level feature aggregation method is proposed, which consists of two parts, i.e., Average Aggregation Strategy (AAS) and Raw Feature Utilization (RFU). The experimental results demonstrate that this method can boost the performance of existing models. In particular, with the help of our clip-level feature aggregation method, we achieve 87.7% rank-1 and 82.3% mAP on MARS dataset without any post-processing procedure, which outperforms the existing state-of-the-art methods

Acknowledgments: This work was supported by EU H2020 MSCA through Project ACHIEVE-ITN (Grant No 765866).